Anomalous payload detection system using MUXConv neural network with parameter optimization

CholRyong Pak*, HakMyong O, HyokChol U, Hun Nam
Faculty of Information Science, **Kim Il Sung** University, Pyongyang,
Democratic People's Republic of Korea

1

***Abstract*— This paper proposes how to detect malicious network data in effective and accurate way using MUXConv neural network(MUXCNN) with parameter optimization. First of all, in order to increase detection speed, packets are directly entered into the input of MUXCNN without decoding. Next of all, after training MUXCNN with learning data, we judge that its traffic is normal or abnormal. Simulations and experiments show that the proposed abnormal network-detecting system is more efficient in detection and higher in accuracy than the other multi-layer neural networks.**

***Keywords*— Network Intrusion Detection System(NIDS), Artificial Neural Network(ANN), Deep Learning(DL), Convolution neural network(CNN), MUXConv, MUXCNN**

I. INTRODUCTION

As more people use the Internet for personal or business reasons, different cyber-attacks and intrusions are growing by the day. IDS is one of among the most essential consideration of cyber-security. IDS is utilized to recognize successful violations even after they have happened [1].

According to Alex Shenfield and his co-authors stated that the research carried out an offline approach for detecting shellcode patterns within a various of file data using artificial neural networks [2].

V. Kanimozhi and his co-authors stated that their proposed system using ANN can be applied to conventional network traffic analysis, cyber-physical system traffic analysis and also can be applied to the real-time network traffic data analysis [3].

In this paper we propose a method to detect anomalous payload using CNN on the network traffic.

Results presented show that this classification approach is capable of detecting anomalous payload with extremely high accuracy and minimal numbers of false positives.

The proposed approach is validated using repeated 10-fold cross-validation and is then tested with respect to creation of false positive alerts on a large dataset of typical network traffic contents (achieving a false positive rate of less than 1%).

The rest of this paper is organized as follows: section II provides a background to intrusion detection systems and convolution neural network, before section III provides a brief introduction to the particular instances that motivated the creation of this system and the results achieved by the proposed MUXCNN based intrusion detection system. Section IV then concludes with the main achievements of this research and for further work.

II. BACKGROUND AND PREVIOUS WORK

A. INTRUSION DETECTION SYSTEMS

The term intrusion detection system was first used by James Anderson [4] in the late 70s and early 80s. He introduced the concept of misuse detection and predefine events and provided the basic for future IDS design and development. An IDS is software or hardware designed to detect any malicious activity or attack against the system or network. An IDS collects data from different sources within a computer or a network such as system command, system log, system accounting, security log and network log. Then, it analyzes them to identify possible security violation, and finally, it issues an alert to the system administrator to deal with the intrusion.

The authors Swathi Pai M, Ashoor et al. [5, 6] summarized IDS Functions as: monitoring and analyzing both the user and system

[1]Corresponding author: CholRyong Pak (email: CL.PAK@star-co.net.kp).

activities, analyzing system configurations and vulnerabilities, evaluating the system and file integrity, recognizing patterns typical of attacks, analyzing of abnormal activity patterns and tracking user policy violations.

There are two main types of IDS: Network-based IDS and Host-based IDS [7].

NIDS is placed along a network to monitor all network traffic [7].

HIDS placed on a host to scan and monitor the all hosts process or devices on the network [7].

IDS can be further categorized into anomaly and signature based systems.

Signature-based systems store attack pattern data in signature database to compare the intruded data and judge when they are identical each other. The advantage is that it has a high efficiency of detecting attacks listed in pattern database, while disadvantage is that it's difficult to cope with unknown attacks or well-modified known attacks. Moreover, it needs an expert to keep updating attack database.

Unlike signature-based systems, the above ones, on the basis of taking statistical feature for normal behavior into account, consider any feature to be anomalous once it is different from normal behavior. However, it is difficult to get statistical feature on normal behavior and it has a high percentage of false positive alert. To be worse, it fails to give an alert on anomalous thing.

B. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network(ANN) is a very powerful tool to deal with signal processing, computer vision and other classification and regression problems [9].

ANN inspired by the biological neural network of human brain, is based on a set of algorithms to extract high-level abstract features from input data by multiple processing layers and it has the ability to automatically infer rules for expected results [10-15].

Deep Neural Network(DNN) which has so many complex neurons and layers endowed the function of feature extractor that can express the human's recognition diagnosis efficiently in its model, is relatively a very complicated ANN [14,15].

DNN is large set of algorithms which has the function of extracting feature for recognition automatically and its architecture is different according to applying object [18].

For example, generally, deep Convolution neural network(CNN) is used in image recognition and Recurrent Neural Network(RNN) is for neural machine translation [18].

Deep Learning can be classified as: Deep Neural Network (DNN), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) and Q-learning [16].

Specially, Convolutional neural networks provide an efficient method to constrain the complexity of feedforward neural networks by weight sharing and restriction to local connections and such a convolution neural networks with local weight sharing topology gained considerable interest both in the field of speech and image analysis [17].

The name "convolutional neural network" indicates that the network employs a mathematical operation called *convolution* which is a specialized kind of linear operation [18].

CNNs are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers [17].

ANNs have also been used in several computer security domains, including the analysis of software design flaws [4] and computer virus detection [18] and malware detection [20].

ANN approaches to detection of multiple types of network attacks have also been shown to be effective [19], though their application to the detection of shellcode was not considered [2].

In this paper we detected normal/abnormal packets which are gained from the packet data that are not undergoing artificial pre-processing, just like human recognize the objects from the natural image.

MUXConv was designed by Zhichao Lu, Kalyanmoy Deb, Vishnu Naresh Boddeti to increase the flow of information by progressively multiplexing channel and spatial information in the network, while mitigating computational complexity. [23]

The multiplexed convolution layer, called MUXConv, is a combination of two components: **spatial multiplexing** which enhances the expressivity and predictive performance of the network, and **channel multiplexing** which aids in reducing the computational complexity of the model. [23]

Experimental results on image classification, object detection and transfer learning suggest that MUXNets are able to match the predictive accuracy and efficiency of current state-of-the-art models while be more compact. [23] .

III. ANOMALOUS PAYLOAD DETECTION IN COMPLEX NETWORK TRAFFIC

A. PROBLEM DOMAIN

In general, attack pattern (for example, Buffer Overflow, SQL Injection) attackers appear is one of packet data division on computer network.

Though this pattern for the above attacks is well-known, well-modified pattern is able to get round intrusion detection system so that attackers can avoid to be easily detected.

For example, due to low level of code, small size and well-modified attacking approaches when detecting shellcode in complex network traffic, network-based intrusion detections face many types of challenges.

This enables signature-based detections to generate false positive alert.

Therefore, developing and researching how to detect intrusion in an intellectual way is one of the main trends.

This paper describes a method to identify anomalous packet by means of low rate of false alert on network using CNN, one of the Deep Neural Networks.

MUXConv [23] was used to detect the anomalous packet.

As compared to the conventional CNN, MUXConv increases more information that flows through the channel by using spatial multiplexing and channel multiplexing to enhance the expressivity of the net and mitigate the computational complexity and compactness of the net.

In this paper, when performing the spatial multiplexing in MUXConv, we get a proper ratio of the number of 3 groups of channels, that is, the ratio of the number of channels subjected to super pixel multiplexing operation to those subjected to sub pixel multiplexing operation to those without subjected to any multiplexing operations, resulting in minimizing the computational complexity and compactness of each layer.

MUXCNN works on the basis of the above principle.

We consider that network traffic is represented as binary image files and both normal traffic and anomalous one can be classified by very low generation rate of false alert using MUXCNN.

Intrusion detection system based on MUXCNN distinguished normal traffic from anomalous one with high accuracy rate.

B. CONVOLUTIONAL NEURAL NETWORKS DESIGN

NAS [23] algorithm to simultaneously optimize compactness, efficiency, and accuracy of MUXCNN designed with MUXConv as the basic building block takes very long time to search the optimal model parameters such as kernel size, expansion ratio and leave-out ratio for setting up the channel multiplexing and the spatial multiplexing.

However, if MUXConv is used without NAS algorithm instead of conventional convolution layer, high accuracy of the anomalous payload detection could be achieved.

From the above facts, this paper proposes MUXCNN and integrates it with an analytic optimization algorithm for searching the optimal super pixel and sub pixel multiplexing operation parameters to detect anomalous payload in the computer network.

Assume that the numbers of channels that are grouped into 3 groups are C1, C2 and C3 in the spatial multiplexing, respectively and the numbers of channels that are output through the group-wise convolution are $C_1^{'}, C_2^{'}$ and $C_3^{'}$ in each group, respectively and the total number of channel is $C^{'}$.

$$C = C_1 + C_2 + C_3 \ , \ \ C^{'} = C_1^{'} + C_2^{'} + C_3^{'}.$$

Where $C_1 = \eta_1 C, C_2 = \eta_2 C$.

Then, $C = \eta_1 C + \eta_2 C + (1 - \eta_1 - \eta_2)C$, where $0 < \eta_1, \eta_2 < 1$.

The paper deals with the computational complexity and the compactness under the assumption which the numbers of output channels are the same as the above relation, $C_1^{'} = \eta_1 C^{'}, C_2^{'} = \eta_2 C^{'}$ where $0 < \eta_1, \eta_2 < 1$.

In case that stride is ignored, the computational complexity is expressed as follows;

$$\begin{aligned} &\frac{W}{r}\frac{H}{r} C_1 C_1^{'} r^2 k^2 + WHk^2 C_2 C_2^{'} + WrHr\frac{C_3}{r^2}k^2 \\ &= WHk^2 C_1 C_1^{'} + WHk^2 C_2 C_2^{'} + WHk^2 C_3 C_3^{'} \\ &= WHk^2 CC^{'}(\eta_1^2 + \eta_2^2 + \eta_3^2) \end{aligned} \quad (1)$$

Then, in Equation 1, Equation 2 is satisfied,

$$\eta_2 = \lambda \eta_1, \eta_1 + \eta_2 + \eta_3 = 1, \lambda > 1.0 \quad (2)$$

If we substitute (1) into (2), the computational complexity can be described as follows;

$$WHk^2 CC^{'}(2\eta_1^2(\lambda^2 + \lambda + 1) - 2\eta_1(\lambda + 1) + 1) \quad (3)$$

Here, in case of $\lambda^2 + \lambda + 1 > 0$, (3) has the unique minimum value and η_1 is given as

$$\eta_1 = \frac{\lambda + 1}{2(\lambda^2 + \lambda + 1)} \quad (4)$$

[Theorem]: if $\lambda = r^2$, compactness becomes minimum value.

[Proof]: The compactness can be described below when the stride is not considered.

$$C_1C_1^{'}k^2r^2+C_2C_2^{'}k^2+\frac{C_3C_3^{'}k^2}{r^2}=k^2\eta_1^2CC^{'}r^2+k^2\eta_2^2CC^{'}$$
$$+\frac{k^2\eta_3^2CC^{'}}{r^2}=k^2CC^{'}(\eta_1^2r^2+\lambda^2\eta_1^2+\frac{(1-\eta_1-\lambda\eta_1)^2}{r^2}) \quad (5)$$

If we differentiate (5) with η_1, the following equation is given as follows;

$$2\eta_1r^2+2\lambda^2\eta_1+\frac{2\eta_1+2\lambda^2\eta_1-2-2\lambda+4\lambda\eta_1}{r^2}=0$$

From the above,

$$\eta_1=\frac{1+\lambda}{\lambda^2(r^2+1)+2\lambda+r^4+1} \quad (6)$$

Equation 7 is given by differentiating (5) with λ.

$$2\lambda\eta_1^2+\frac{2\lambda\eta_1^2-2\eta_1+2\eta_1^2}{r^2}=0 \quad (7)$$

From this,

$$\lambda=\frac{1-\eta_1}{\eta_1r^2+\eta_1} \quad (8)$$

If we substitute Equation 6 into Equation 8, Equation 9 is given as,

$$\lambda=r^2 \quad (9)$$

Here, r is a natural number that satisfies the following Equation 10, (in case of $W=H$)

$$\begin{cases} W \bmod r=0 \\ (1-\eta_1-\lambda\eta_1)C \bmod r^2=0 \end{cases} \quad (10)$$

The model consists of the part for composing feature map using VGG19[21], the T($T\geq 2$)stages for extracting confidence map and the classifier.

LeakyRelu is used in all convolution layers as activation function and Global average pooling and Softmax are used in the last layer of the classifier.

The definition of LeakyRelu is shown below.

$$leakyReLu=\alpha x,\ for\ x<0$$
$$leakyReLu=\max(x,0),\ for\ x\geq 0 \quad (11)$$

Here, α is user-defined parameter and its value is about 0.3-04.

The structure of feature extractor is given in table 2.

A stage contains 3 convolution layers which have 64 filters of size 1*1, 64 filters of size 3*3 and 256 filters of size 1*1, respectively.

Classifier contains three convolution layers which have 64 filters of size 3*3, 32 filters of size 3*3 and 2 filters of size 3*3 respectively, along with a global average pooling layer and Softmax layer.

For experiment, byte data of payload of traffic collected are input into MUXCNN.

Data for training and testing of MUXCNN are collected by network traffic for 5 days.

Table 1 shows the description of dataset.

Lack of the length of data division to be input is added by zero padding, which is input into convolution layer as a type of input image.

Output layer has 2 neurons to classify normal and anomalous things.

Day & property	Contents
Monday / 14G	Normal Traffic
Tuesday / 11G	Attack Traffic + Normal Traffic
Wednesday / 13G	Attack Traffic + Normal Traffic
Thursday / 13G	Attack Traffic + Normal Traffic
Friday / 11.5G	Attack Traffic + Normal Traffic

Anomalous data come from Kali and Acunetix Web Vulnerability Scanner

The total dataset consists of three parts - training dataset, validation dataset and test dataset.

Training dataset consists of 300000 normal packet data, 250000 abnormal packet data.

Validation dataset consists of 300000 normal packet data and 250000 abnormal packet

data, of which trained packet data is 70% and unknown data is 30%.

Test dataset is also unknown dataset, which contains 50000 normal packets and 50000 abnormal packets.

Figure 1 shows the structure of MUXCNN model.

TensorFlow in “Python programming tutorials” is used to classify anomalous traffic.

Computer CPU used in experiment is XEON E3 1210 v6 and its hard disk is RAID 5, 4 TB.

.

C. Results

ROC curve is used to visualize multi-dimensional performance.

Figure 2 shows ROC result.

AUC mark is 0.99989.

In our experiment, we set T as 3.

Classification accuracy for training dataset is 1.0 and that for validation dataset is 0.999187 and that for testing dataset is 0.98248.

Name	Type	Input & output	Description
Data	Input	$1\times40\times40$	Image data input
conv1	convolution	$64\times40\times40$	kernel:3, stride:1
conv2	convolution	$64\times40\times40$	kernel:3, stride:1
pool1	Max pooling	$64\times20\times20$	kernel:3, stride:2
conv3	convolution	$128\times20\times20$	kernel:3, stride:1
conv4	convolution	$128\times20\times20$	kernel:3, stride:1
pool2	Max pooling	$128\times10\times10$	kernel:3, stride:2
conv5	convolution	$256\times10\times10$	kernel:3, stride:1
conv6	convolution	$256\times10\times10$	kernel:3, stride:1

Table 2. Structure of each layer of CNN

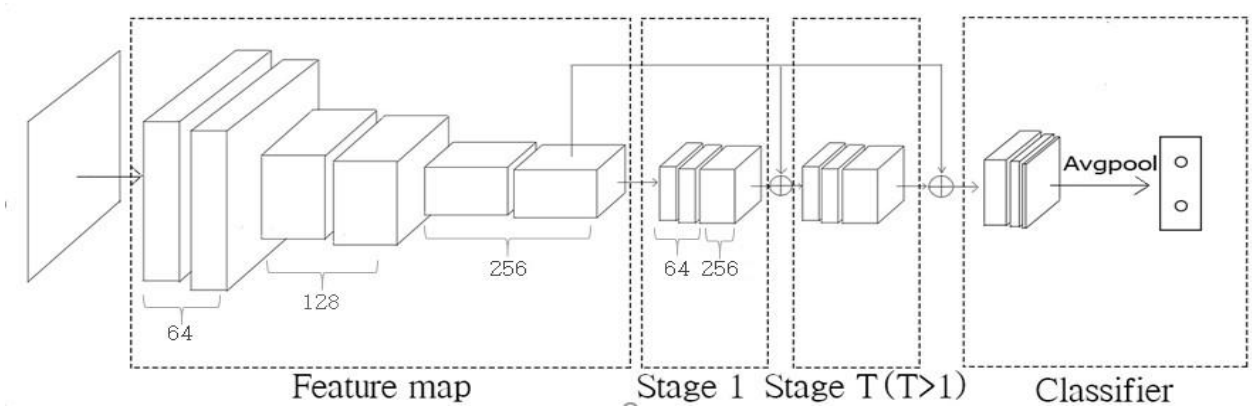

Figure 1. MUXCNN model for classification of anomalous traffic

For comparison, a general deep neural network is used with 1 input layer, 100 neurons, 3 intermediate layers, 100 neurons of each layer and 2 neurons of output layer.

We also used DL4MD [20], Support Vector Machine (SVM), VGG-16[21], SENet[22] for comparison.

SENet is the model that MUXConvs in the MUXCNN are replaced by SE blocks.

Fully connected layer with 2 neurons is added to the last layer of VGG-16 to classify normal/anomalous payload.

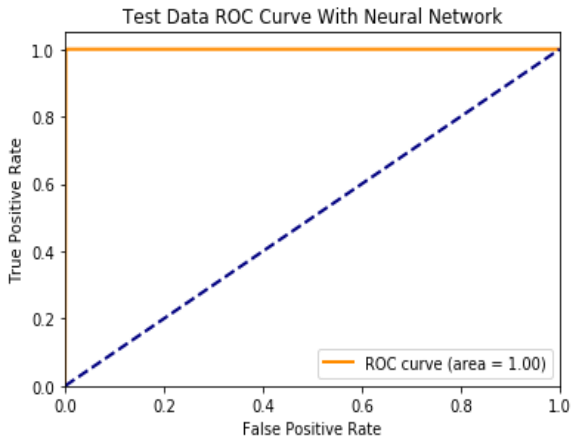

Figure 2. ROC result

Method	TP	FP	TN	FN	ACY
ANN	248647	2603	297397	1353	0.992807
DL4MD	249437	1169	298831	563	0.996851
SVM	248032	3822	296178	1968	0.989473
SENet	249998	0	300000	2	0.999996364
VGG-16	250000	0	300000	0	1.0
MUXCNN	250000	0	300000	0	1.0

Table 3. Training Dataset

Method	TP	FP	TN	FN	ACY
ANN	237423	20604	279396	12577	0.939671
DL4MD	240956	17856	282144	9044	0.951091
SVM	236085	22562	277438	13915	0.933678
SENet	248641	1037	298963	1359	0.995643636
VGG-16	247969	1453	298547	2031	0.993665455
MUXCNN	249790	237	299763	210	0.999187

Table 4. Validation Dataset

Method	TP	FP	TN	FN	ACY
ANN	42413	8732	41268	7587	0.83681
DL4MD	42791	6316	43684	7209	0.86475
SVM	39785	9880	40120	10215	0.79905
SENet	48917	680	49320	1083	0.98237
VGG-16	48319	1790	48210	1681	0.96529
MUXCNN	48951	703	49297	1049	0.98248

Table 5. Testing Dataset

IV. Conclusion and Future Work

Intellectual intrusion detecting system mentioned in this paper improves the detecting method of anomalous intrusion using MUXCNN to classify normal traffic and anomalous one in network traffic.

Classifier using MUXCNN is not only fully sensitive to checking dataset but also high in accuracy.

This paper describes offline method to classify normal and anomalous traffic within data.

This method is currently in use in online network intrusion detection based on anomalous detection of hardware base like FPGA.

Future work is to classify anomalous traffic by deep learning to subdivide them into some typical types of attack.

V. References

[1] Akbar, S., T.S. Rao, and M.A. Hussain, A Hybrid Scheme based on Big Data Analytics using Intrusion Detection System. Indian Journal of Science and Technology, 2016. 9(33).

[2] Alex Shenfield, David Day, and Aladdin Ayesh, "Intelligent intrusion detection system using artificial neural networks," ICT Express, June 2018.

[3] V. Kanimozhi and T.P. Jacob, " Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing", ICT Express, 2019.

[4] Anderson, J.P., Computer security threat monitoring and surveillance. 1980, Technical report, James P. Anderson Company, Fort Washington, Pennsylvania.

[5] Swathi Pai M., B.B.K., Big Data Security Analytic: A classification technique for Intrusion Detection System. ResearchGate, 2015.

[6] Ashoor, A.S. and S. Gore, Importance of intrusion detection system (IDS). International Journal of Scientific and Engineering Research, 2011. 2(1): p. 1-4.

[7] Soniya, S.S. and S.M.C. Vigila. Intrusion detection system: Classification and techniques. in Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on. 2016. IEEE.

[8] S. Hou, L. Chen, E. Tas, I. Demihovskiy, and Y. Ye. "Cluster-Oriented Ensemble Classifiers for Malware Detection." In IEEE International Conference on Sematic Computing (IEEE ICSC), 189-196, 2015.

[9] Shuai Li, Ken Choi and Yunsik Lee, Artificial Neural Network Implementation in FPGA: A Case Study, ISOCC, 297-298, 2016.

[10] L.Deng, D. Yu, "Deep Learning : Methods and Applications", Foundations and Trends in Signal Processing 7:3-4,2014

[11] Y.Bengio,"Learning Deep Architectures for AI", Foundations and Trends in Machine Learning 2(1):1-127,2009.

[12] Y. Bengio, A. Courville, P.Vincent, "Representation Learning: A Review and New Perspectives", IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8): 1798-1828,2013.

[13] J. Schmidhuber, "Deep Learning in Neural Networks": An Overview", Neural Networks 61: 85-117,2015

[14] Y.Bengio, Y. LeCun, G. Hinton, "Deep Learning", Nature 521: 436-444,2015

[15] I. Arel, D. C. Rose, T. P. Karnowski, "Deep Machine Learning"- A New Frontier in Artificial Intelligence Research", IEEE Computational Intelligence Magazine, 2013.

[16] Bendangnuksung, Dr. Prabu P," Students' Performance Prediction Using Deep Neural Network", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2 (2018) pp. 1171-1176

[17] Claus Neubauer, "Evaluation of Convolutional Neural Networks for Visual Recognition", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 9, NO. 4 , 1998

[18] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning", 2016, MIT Press: Cambridge, MA.

[19] J. M. Vidal, A. L. S. Orozco, and L. J. G. Villalba. Quantitative criteria for alert correlation of anomalies-based nids. IEEE Latin America Transactions, 13(10):3461–3466, 2015.

[20]. William Hardy, Lingwei Chen, Shifu Hou, Yanfang Ye, and Xin Li, DL4MD: A Deep Learning Framework for Intelligent Malware Detection, Int'l Conf. Data Mining:61-67, CSREA Press, 2016.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[22] J. Hu, Li. Shen, S. Albanie, G. Sun and E.Wu, " Squeeze-and-Excitation Networks", *arXiv preprint arXiv:1709.01507v4*, 2019.

[23] Z. Lu, K. Deb, V. N. Boddeti, "MUXConv: Information Multiplexing in Convolutional Neural Networks" , in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12 044–12053.